



k-匿名化

TF-IDF

匿名加工情報

個人情報の匿名化技術

河野 和宏

社会安全学部 安全マネジメント学科

Point1

本研究の概要

匿名化で重視すべきポイントは、個人に関する情報の価値を維持しながらプライバシーを守ることです。特定の個人の識別を困難にするためのデータ加工方法のひとつ。「k-匿名化」とは、対象となるデータ内に、同じ属性値を持つデータがk件以上存在するようにデータを変換することで、個人が特定される確率をk分の1以下に低減させることができます。本技術では、匿名化によるデータの情報量の損失の少なさよりも、サービス提供者が推薦したい情報の属性のデータが残るよう、TF-IDF法を用いて属性を重みづけして匿名化されたデータを生成します。

Point2

応用可能な分野

2022年施行改正個人情報保護法によって、イノベーションを促進する観点からデータ利活用の促進個人を特定できないように個人情報を加工した場合は（仮名加工情報）、仮名加工情報取扱事業者の内部分析目的の利用に限定する等を条件に、開示・利用停止請求への対応等の義務が緩和されました。ビッグデータの利活用が進む一方で、事業者の責務拡大とペナルティ強化が図られています。

Point3

連携を希望する業種等

上記、応用可能な分野に関連すれば業種は問いません。

詳細な研究・技術シーズは次のページへ



個人情報 の 匿名化 技術

用途・応用分野

- ・ 個人情報保護法で求められている匿名加工情報の作成
- ・ プライバシーを保護しつつ個人にあった情報推薦システムサービス

本技術の特徴・従来技術との比較

- ・ 個人情報保護法で求められている匿名加工情報(k -匿名化された情報)の生成を実現
- ・ k -匿名化に利用されるアルゴリズムでは、匿名化された際になるべく情報を残すように匿名化するが、本技術では、サービス提供者ごとに必要な情報は異なるという観点から、サービス提供者が情報提供(推薦)したい情報に基づいて、個人情報を匿名化する
- ・ サービス提供者ごとにあった匿名化された個人情報の提供が可能

技術の概要

○ k -匿名化された個人情報とは

k 人以上が同じ属性値の組み合わせを持ち、区別できない状態に変換

○本技術での k -匿名化された個人情報の生成法

匿名化によるデータの情報の損失の少なさよりも、サービス提供者が推薦したい情報の属性のデータが残るよう、TF-IDF法を用いて属性を重みづけして匿名化されたデータを生成

疑似個人情報のデータ

ID	氏名	性別	生年月日	職業	身長
1	坂田 聡	男性	1986年8月13日	アナウンサー	188cm
2	荒川 礼子	女性	1993年11月25日	高校生	165cm
3	山田 舞子	女性	1968年11月9日	パティシエ	160cm
⋮	⋮	⋮	⋮	⋮	⋮

各店舗(サービス提供者)における推薦情報と必要な属性

RecNo	推薦情報	性別	生年月日	職業	身長
店1	ネクタイ	男性	*	*	*
店2	服A	*	*	*	160cm代

ネクタイを推薦した場合の匿名化の結果



性別	生年月日	職業	身長	人数
男女	10年毎	*	*	39
男女	*	*	10cm毎	53
男女	10年毎	*	10cm毎	8

服Aを推薦した場合の匿名化の結果



性別	生年月日	職業	身長	人数
男女	*	*	10cm毎	73
*	10年毎	*	10cm毎	12
男女	10年毎	*	10cm毎	15

特許・論文

1. K. Kono and N. Babaguchi, "Data Anonymization for Service Strategy Development and Information Recommendation to Users Based on TF-IDF Method," Proc. of 30th International Symposium on Information Theory and Its Applications (ISITA 2020), pp. 485-489, Hawaii, USA (Virtual Conference), October 2020.
2. 新井健介, 河野和宏, 馬場口登, "推薦対象の属性から構築した階層構造を用いたTF-IDF法による匿名化処理," 電子情報通信学会技術研究報告, vol. 115, no. 479, EMM2015-81, pp. 31-36, 2016.

研究者

河野 和宏

社会安全学部 安全マネジメント学科
河野研究室

